‘Corpusgebaseerd onderzoek naar morfologische produktiviteit’

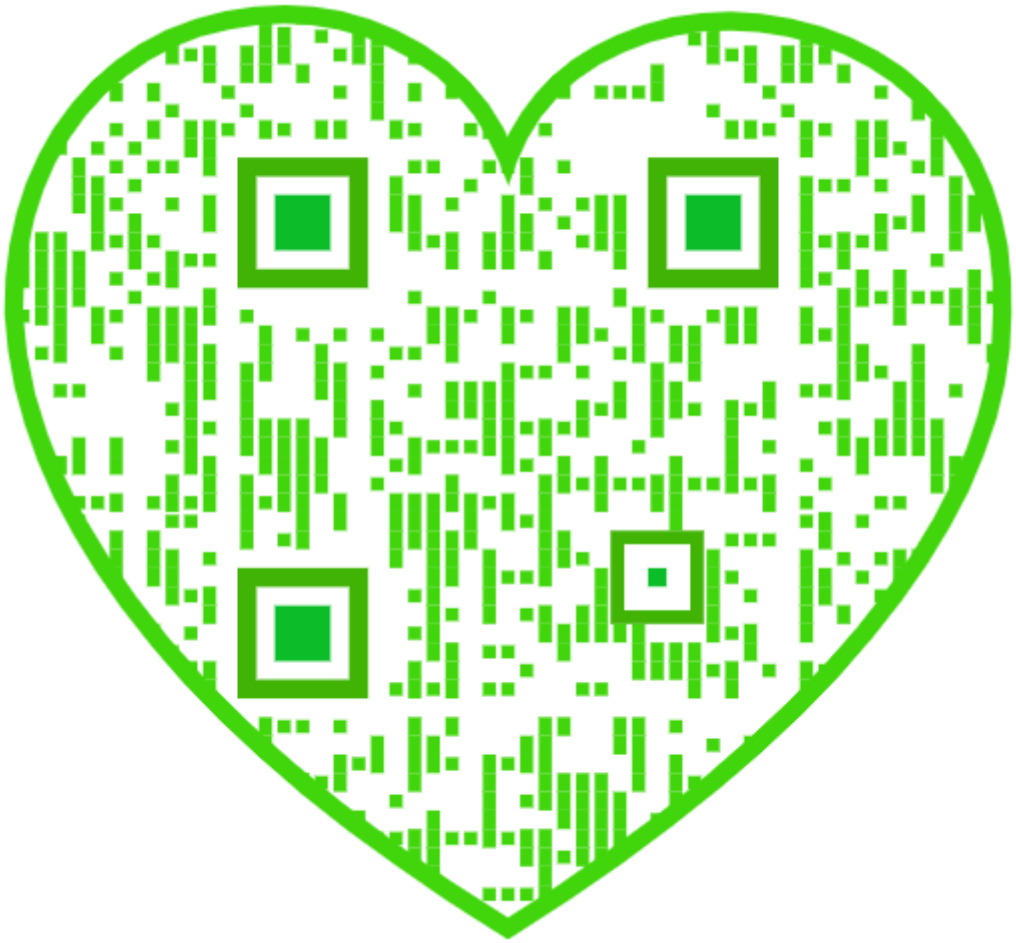
R.H. Baayen

bron
R.H. Baayen, ‘Corpusgebaseerd onderzoek naar morfologische produktiviteit.’ In: *Spektator* 19 (1990), p. 213-233.

Zie voor verantwoording: https://www.dbnl.org/tekst/baay001corp01_01/colofon.php

Let op: boeken en tijdschriftjaargangen die korter dan 140 jaar geleden verschenen zijn, kunnen auteursrechtelijk beschermd zijn. Welke vormen van gebruik zijn toegestaan voor dit werk of delen ervan, lees je in de gebruiksvoorwaarden.

dbnl digitale bibliotheek voor de Nederlandse letteren



Corpusgebaseerd onderzoek naar morfologische produktiviteit

R.H. Baayen

1. Inleiding

Intuïties ten aanzien van de spontane uitbreidbaarheid van morfologische categorieën spelen een zeer belangrijke rol bij het bepalen van produktiviteit. * In het licht van de belangwekkende resultaten die door een zorgvuldig onderzoek van wat intuïtief als mogelijk wordt ervaren zijn bereikt, lijkt op het eerste gezicht een corpusgebaseerde, kwantitatieve benadering weinig nieuwe inzichten te kunnen bieden. Dat een statistische analyse van de frekwenties waarmee de soorten in het corpus voorkomen wel degelijk vruchtbaar kan zijn, en zelfs tot nieuwe inzichten kan leiden, hoop ik in dit artikel duidelijk te maken.

Het nut van statistisch onderzoek van het kwantitatieve aspect van morfologische produktiviteit laat zich goed illustreren aan de hand van het probleem van divergerende intuïties over de produktiviteit van morfologische categorieën. Volgens Schultink (1962) is het suffix *-te* improduktief. Zijns inziens zijn formaties als *kwaadte* en *mooite* onwelgevormd. Daarentegen stelt Booij (1977) dat *-te* niet improduktief is, zij het dat het slechts een geringe graad van produktiviteit kent. Volgens hem zijn formaties als *geelte* en *gekte* wel degelijk grammaticaal, al hebben ze voor sommige sprekers wellicht een zeer lage graad van acceptabiliteit. Voor deze opvatting pleit het feit dat een woord als 'gekte' zich inderdaad lijkt in te burgeren in onze taal. Toch hebben sommige sprekers er nog moeite mee, hetgeen blijkt uit de aanhalingstekens rond dit woord in een artikel uit Trouw van 21-1-1987:

> De Haagse bewonersorganisatie …beschouwt de plannen van Den Haag voor een nieuw stadhuis als 'gekte'.

Echter, de centrale vraag die hier gesteld moet worden is of het in het geval van het suffix *-te* mogelijk is om onopzettelijk een in principe aftelbaar oneindig aantal nieuwe formaties te vormen. We zullen zien dat een statistische analyse van ons materiaal uitwijst dat dit bijzonder onwaarschijnlijk is. Een kwantitatieve analyse werpt ook enig licht op een discussie over de produktiviteit van het suffix *-atie*. Van Marle (1985) stelt dat in het niet inheemse stratum van het Nederlandse lexicon geen produktieve processen worden gevonden. Hij spreekt hier liever van 'creative coining'. Booij (1977), daarentegen, ziet geen reden om de produktiviteit van *-atie* in twijfel te trekken. Een statistische analyse, waarin ook het rivaliserende affix *-ing* betrokken wordt voor zover het zich hecht aan woorden op *-eer*, voortaan aan te duiden als *-ering*, wijst uit dat de produktiviteit van beide processen een enigszins ambivalent karakter vertoont. In het licht van dit ambivalente karakter blijken de op het eerste gezicht incompatibele intuïties van van

Marle en Booij complementaire aspecten van de produktiviteit van deze suffixen te betreffen.

Tenslotte heeft een kwantitatieve analyse een heuristische waarde, die in sommige gevallen ondersteunende evidentie aandraagt, en in andere een heranalyse noodzakelijk maakt. Dit aspect van de hier voorgestelde kwantitatieve methode illustreer ik aan de hand van een aantal verbale morfologische categorieën.

Als ik in dit artikel mij vrijwel uitsluitend op het kwantitatieve aspect van morfologische produktiviteit richt, wil daarmee niet gezegd zijn dat produktiviteit louter frekwentieel bepaald is. De relevantie van beperkende factoren (Booij 1977), van de paradigmatiek van rivaliserende affixen (van Marle 1985), van blocking (Rainer 1988) en van semantische transparantheid (van Marle 1988) wordt volledig onderschreven. Op al deze gebieden is en kwantitatieve uitsplitsing mogelijk. Hier beperk ik me echter tot de analyse van de algehele produktiviteit van de morfologische categorie als produkt van de interactie van syntagmatiek, paradigmatiek, blocking, semantiek en ook pragmatiek.

2. Het kwantitatieve aspect in eerder onderzoek

Binnen de Nederlandse morfologie is aan het kwantitatieve aspect van morfologische produktiviteit aandacht besteed door Schultink (1962) en Al & Booij (1981). Terwijl in de studie van Schultink de in de lexica aangetroffen of uit mondeling taalgebruik opgetekende formaties nauwkeurig worden vermeld, waaruit een zekere aandacht voor het kwantitatieve aspect van het aantal soorten (typen) blijkt, wordt hier geen poging gedaan om de produktiviteit te kwantificeren in termen van het aantal geregistreerde soorten (typefrekwentie). Een dergelijke kwantificatie gaat namelijk voorbij aan het inzicht dat een improduktieve categorie een gering en dus eindig aantal soorten bevat, terwijl een produktieve categorie een aftelbaar oneindige verzameling soorten bevat. Naast dit door Schultink uitvoerig gemotiveerde principiële bezwaar is er eveneens een practisch probleem, dat zich voordoet zodra men produktiviteit tracht te meten in termen van het aantal soorten. Bijvoorbeeld, de door Schultink als improduktief gekarakteriseerde morfologische categorie van de-adjectivale formaties met het suffix -*te* telt rond de negentig soorten (Schultink 1962: 178-180). De categorie van de-adjectivale nomina met het suffix -*erd* telt daarentegen slechts een vijftigtal soorten (Schultink 1962: 200-202). Dat het suffix -*te* ondanks het grotere aantal soorten, toch als improduktief wordt gekarakteriseerd, vindt zijn motivatie in het feit dat nieuwvormingen in -*te* niet onopzettelijk gevormd kunnen worden (Schultink 1962: 180). Omgekeerd beschouwt Schultink het suffix -*erd* als produktief in het licht van 'het gemak waarmee deze formaties gevormd kunnen worden' (Schultink 1962: 202). Dat een aantal van de door hem opgetekende formaties in -*erd* niettemin een opzettelijk karakter lijken te vertonen ziet Schultink niet als een bezwaar, aangezien opzettelijk gebruik de mogelijkheid van spontane vorming niet uitsluit.

De intuïtie van de taalgebruiker is in deze benadering het centrale criterium voor de produktiviteit van morfologische categorieën. Vooropgesteld dat een

zinvolle kwantitatieve maat voor de (graad van) produktiviteit van een morfologische categorie aan de intuïties van de taalgebruiker ten aanzien van de produktiviteit van de categorie in kwestie recht dient te doen, moeten we toch vaststellen dat een alleen van intuïties gebruik makende morfologie met name in grensgevallen zich voor nauwelijks oplosbare problemen gesteld ziet. In de inleiding heb ik reeds het probleem genoemd dat in dit soort situaties intuïties kunnen divergeren. Booij (1977) en Schultink (1962) schatten bijvoorbeeld de produktiviteit van *-te* verschillend in. Maar ook bij het suffix *-erd* kan men zich afvragen of hier wel van produktiviteit sprake is. Het feit dat een aantal formaties in *-erd* kennelijk opzettelijk is gevormd leidt niet noodzakelijk tot een negatief productiviteitsoordeel, maar is anderzijds ook niet met een dergelijk negatief oordeel in strijd. Ten aanzien van formaties als

(1)
luttelerd
paarserd
stukkerd
zilterd

valt het te betwijfelen of ze met dezelfde vanzelfsprekendheid als welgevormd worden ervaren als bijvoorbeeld de volgende formaties met het suffix *-heid*: [1]

(2)
ontegenzeggelijkheid
afschrijfbaarheid
bewandelbaarheid
doortastendheid

Een preciese en in grensgevallen objectieve maat voor de (graad van) produktiviteit die recht doet aan de intuïties in de duidelijke gevallen kan hier verhelderend werken.

Nu is het mogelijk een dergelijke maat voor de graad van produktiviteit te ontwikkelen zodra de gebruiksfrekwenties van de soorten (de tokenfrekwenties) in de analyse betrokken worden. De relevantie van deze gebruiksfrekwenties lijkt op het eerste gezicht twijfelachtig. Op grond van een tweetal Engelstalige frekwentiestudies concludeerde Schultink (1962: 37) dat ‘op zichzelf belangrijk woordfrekwentieonderzoek ons over de produktiviteit van categorieën maar bitter weinig informatie verschaft’. Nu geven de studies van Harwood & Wright (1956) en Thorndike (1943), om redenen die in Baayen (1989) uitvoerig worden toegelicht, inderdaad weinig zicht op de rol van gebruiksfrekwenties. Het blijkt echter dat, zoals hoge gebruiksfrekwenties karakteristiek zijn voor duidelijk improduktieve categorieën - vergelijk Schultink (1962: 37) - juist de zeer lage gebruiksfrekwenties kenmerkend zijn voor de produktieve categorieën. Met name het aantal soorten met de laagst mogelijk gebruiksfrekwentie, de gebruiksfrekwentie 1 (hapaxen), speelt een cruciale rol bij het bepalen van ‘the statistically determinable readiness with which an element enters into new combinations’ (Bolinger 1948: 18). Voordat we meer in detail de op dit inzicht gebaseerde maat voor de graad van produktiviteit introduceren, behandelen we eerst de door Al & Booij (1981) voorgestelde kwantificering van morfologische produktiviteit.

Al & Booij (1981) kiezen hun uitgangspunt bij de in Aronoff (1976) in overweging genomen produktiviteitsindex I = V/S, met *V* het aantal geattesteerde soorten en *S* het aantal mogelijke soorten. Door paren rivaliserende affixen met elkaar te vergelijken ontstaat de mogelijkheid om een relatieve toepassingswaarschijnlijkheid in te voeren waarin het aantal mogelijke soorten *S* geen rol meer speelt. Immers, rivaliserende affixen hebben een gemeenschappelijk derivationeel domein, zodat we kunnen overgaan tot het vergelijken van V_1 en V_2, de aantallen soorten van beide affixen. De relatieve toepassingswaarschijnlijkheden van de beide rivaliserende affixen kunnen nu geïntroduceerd worden als

$$P_{rel}(WFR1, WFR2) = V_1/(V_1 + V_2)$$

$$P_{rel}(WFR2, WFR1) = V_2/(V_1 + V_2)$$

Vervolgens splitsen zij de V_i uit naar de aantallen soorten die vallen onder de verschillende beperkende factoren (competence restrictions) die voor het suffixpaar van kracht zijn. Op de uitgesplitste relatieve frekwenties passen zij het multiplicatieve model van beperkende factoren van Cedergren & Sankoff (1974) toe, om te komen tot een model waarin het gewicht van de verschillende beperkende factoren aan de toepassingswaarschijnlijkheid van de regel als geheel kan worden afgelezen.

De aantrekkelijkheid van deze benadering ligt in het feit dat het hier gaat om een methode van onderzoek die, uitgaande van de frekwenties van de soorten in corpora, [2] de paradigmatiek van rivaliserende affixen per beperkende factor in kaart tracht te brengen.

Helaas reikt dit model een produktiviteitsmaat aan die het improduktieve affix van het ene rivaliserende affixpaar hoger inschaalt dan een produktief affix van een ander rivaliserend affixpaar. Uit een telling van het aantal soorten in *-ering* en *-atie* in het corpus Uit den Boogaart (1975), geschreven taal (voortaan UdB), blijkt dat het aantal soorten waarvoor geen beperkende factoren in het spel is, respectievelijk 46 eh 161 bedraagt. [3] Dit leidt tot relatieve toepassingswaarschijnlijkheden van 0.222 en 0.778. Een telling van soorten in *-te* en *-heid*, waarbij we ons eveneens beperken tot die soorten waarvoor geen beperkende factoren van kracht zijn, [4] resulteert in de respectievelijke aantallen 39 en 90 en de toepassingswaarschijnlijkheden 0.302 en 0.698. We zien nu dat het produktieve *-ering* op een produktiviteitsschaal van o tot 1 wordt gelocaliseerd bij 0.222, terwijl het improduktieve *-te* op dezelfde schaal hoger gevonden wordt, namelijk bij 0.302. Ondanks het feit dat het verschil tussen de toepassingswaarschijnlijkheden van *-te* en *-ering* niet significant is (Z = 1.514), is de produktiviteitsmaat voor *-te* onbevredigend ten opzichte van die voor *-ering*.

Tabel 1. Voorwaardelijke relatieve typefrekwenties van -*ering* en -*atie* in UdB

	0	**1**	**2**	**2 & 3**
-*atie*	161/207	10/30	11/20	2/10
-*ering*	46/207	20/30	9/20	8/10

Nu kunnen we proberen deze aanpak te redden door per beperkende factor het aantal mogelijke soorten te benaderen door het aantal in het corpus aanwezige grondwoorden van de vereiste vorm. Immers, dit aantal grondwoorden geeft een indicatie van het aantal gelede formaties dat zich op basis van wat het corpus aanreikt direkt laat vormen. Voor rivaliserende affixen is dit aantal mogelijke grondwoorden gelijk. Wellicht zouden met deze corpusgebaseerde schatting van *S* de frekwentieverhoudingen zodanig kunnen verschuiven dat -*te* lager wordt ingeschaald dan -*ering*. Een dergelijke verschuiving treedt inderdaad op. Tabel 1 komt er dan uit te zien als tabel 2, terwijl de relatieve toepassingswaarschijnlijkheid van -*te* nu de waarde 39/294 = 0.133 aanneemt. Toch blijft het verschil tussen de beide toepassingswaarschijnlijkheden niet significant ($Z = 0.343$). Een objectieve maat voor de graad van produktiviteit, een maat die een vergelijking mogelijk maakt tussen paren van rivaliserende affixen en die in overeenstemming is met de intuïties van de taalgebruiker, moet blijkbaar op een andere manier worden verkregen.

Tabel 2. Onvoorwaardelijke relatieve typefrekwenties van -*ering* en -*atie* in UdB

	0	**1**	**2**	**2 & 3**
-*atie*	161/316 = 0.509	10/79	11/74	2/26
-*ering*	46/316 = 0.146	20/79	9/74	8/26

3. Maten voor produktiviteit

Om de kwantitatieve aspecten van morfologische produktiviteit in de greep te krijgen ontleden we het begrip produktiviteit in een drietal factoren, te weten pragmatische bruikbaarheid, produktiviteit in engere zin, en pragmatische potentialiteit.

• *Pragmatische bruikbaarheid* (*U*). Hieronder versta ik de inzetbaarheid van de morfologische categorie. Vergelijk bijvoorbeeld de aantallen soorten in UdB van de volgende zes nominale categorieën:

-erd	6
-ster	30
-sel	44
heid	466
-je (en allomorfen)	1031
N + N samenstellingen	4277

Alle zes categorieën staan in de literatuur bekend als produktief. Toch is er een overduidelijk verschil in de mate waarin deze categorieën worden ingezet binnen de 600.000 tokens van het UdB. Samenstellingen lenen zich beter voor frekwent gebruik dan persoonsaanduidende nomina in *-erd*, wellicht dankzij hun semantische wendbaarheid (vergelijk Downing 1977), die het mogelijk maakt ze te gebruiken voor incidenteel te benoemen of uit te duiden objecten. [5] Daartegenover veronderstelt het inzetten van een persoonsaanduidend nomen in *-erd* een specifieke (speelse) gebruikscontext die men eerder in het mondelinge dan in het schriftelijke taalverkeer zal tegenkomen. Met andere woorden, nomina in *-erd* in geschreven taal zijn niet ongrammatikaal, maar hun gebruik in geschreven taal kan minder geschikt zijn dan hun gebruik in gesproken taal. [6] Evenzo geldt dat entiteiten waarvoor een conceptualisatie met behulp van het suffix *-sel* geboden is minder vaak in onze ervaringswereld worden aangetroffen dan die entiteiten die zich met behulp van een samenstelling laten beschrijven. Met de term pragmatische bruikbaarheid bedoelen we dit type verschillen in practische inzetbaarheid aan te geven. De pragmatische bruikbaarheid U meten we in termen van het aantal soorten V (typen) gerelateerd aan het totaal aantal woordvoorkomens in het corpus. Met andere woorden, $V = U^{(F)}$, met dien verstande dat de notatie $U^{(F)}$ benadrukt dat het aantal soorten V binnen een corpus van omvang F is aangetroffen. Gezien het feit dat al de hier besproken gegevens aan het corpus Uit den Boogaart (1977) zijn ontleend, zijn $U^{(F)}$ en V inwisselbaar. (Voor tellingen van het aantal soorten in corpora van verschillende omvang F is daarentegen het vergelijken van V zonder een correctie voor F niet zinvol).

• *Productiviteit in engere zin* (P): de kans dat de woordvormingsregel wordt ingezet, oftewel, de graad van produktiviteit. Mathematisch wordt deze kans zeer goed benaderd door het quotiënt van het aantal hapaxen (n_1) en het totale aantal tokens van de woorden voortgebracht door de regel (N): $P = n_1/N$ (zie Baayen 1989, hoofdstuk 5). Hoe groter het aantal hapaxen n_1, des te groter is de graad van produktiviteit P. Nu bevatten produktieve categorieën in het algemeen tevens relatief weinig soorten met hoge gebruiksfrekwenties. Bijgevolg blijft N relatief laag, hetgeen eveneens een hoge waarde van P bevordert. Omgekeerd kenmerken categorieën met een lage graad van produktiviteit zich door een gering aantal hapaxen en een groot aantal soorten met hoge gebruiksfrekwenties, en dus door een laag quotiënt n_1/N.

• Pragmatische potentialiteit (I): De mate waarin het aantal pragmatisch-realistisch mogelijke soorten (S) door het aantal soorten in de steekproef (V) wordt gerealiseerd. Door de schat aan informatie die besloten ligt in het volledige frekwentiespectrum te exploiteren kunnen we het aantal soorten in de populatie, S, schatten. Daarbij maken we gebruik van een wiskundig model dat generaliseert over de wet van Zipf (zie Baayen 1989, hoofdstuk 6). Met een schatter van S in handen kunnen we nagaan of een categorie produktief is door te onderzoeken of S

oneindig is. Zo ja, dan wordt $I = S/V$ gelijk aan ∞, zoals in het geval van het diminutiefsuffix *-je*. Indien niet, dan wordt de mate waarin de steekproef het aantal soorten in de populatie uitput kleiner naarmate I groter wordt. [7] Improduktieve categorieën kenmerken zich door een lage waarde van I. Bijvoorbeeld, voor het improduktieve suffix *-nis* tellen we 23 soorten in het corpus, terwijl het geschatte aantal mogelijke soorten gelijk is aan 32, hetgeen resulteert in $I = 1.4$. Ook produktieve categorieën kunnen een eindige waarde voor I aannemen, zij het dat deze waarde veel hoger ligt dan het geval is voor improduktieve categorieën. We moeten namelijk rekening houden met het feit dat de aanname dat produktieve woordvormingsregels aftelbaar oneindig veel soorten projecteren tot stand komt binnen een competence theorie, en als zodanig niet noodzakelijk overdraagbaar is op een corpus. Immers, performance factoren hebben bij het tot stand komen van de in het corpus geregistreerde taalbouwsels een rol gespeeld. Voor produktieve woordvormingsprocessen houden we daarom de mogelijkheid open dat het aantal door de performance getemperde, pragmatisch-realistisch mogelijke soorten niet oneindig is. Wel stellen we de voorwaarde dat, wil er van produktiviteit sprake zijn, het aantal mogelijke soorten S substantieel groter moet zijn dan het in de steekproef aangetroffen aantal soorten V. Een voorbeeld van een hoge, maar niet oneindige waarde van I is het suffix *-heid*, waarvoor we vinden dat $I = 2063/466 = 4.4$.

Dat de hier geïntroduceerde maten voor produktiviteit voldoen aan de eisen die elke linguïstische theorie aan dergelijke maten mag stellen, mag blijken uit de nu te behandelen voorbeelden.

4. Pragmatische bruikbaarheid en graad van produktiviteit

In deze paragraaf behandelen we pragmatische bruikbaarheid en graad van produktiviteit van een aantal nominale en verbale morfologische categorieën. We kiezen ons uitgangspunt bij de nominale categorieën, omdat zich hier de meest uitgesproken verschillen in produktiviteit voordoen. Vervolgens illustreren we de mogelijkheden van de hier ontwikkelde kwantitatieve benadering aan de hand van een aantal verbale categorieën.

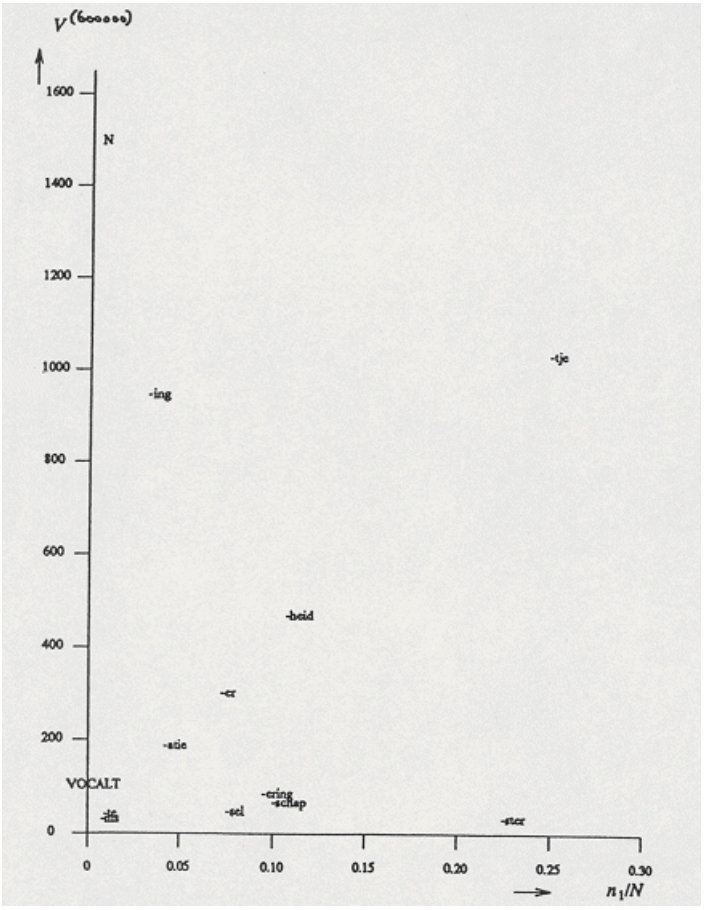
4.1 Nomina

Figuur 1 zet tegen elkaar uit de pragmatische bruikbaarheid op de vertikale as en de graad van produktiviteit op de horizontale as. Niet opgenomen in figuur 1 is de categorie van N + N samenstellingen, in het UdB vertegenwoordigd door 4277 soorten, en een graad van produktiviteit gelijk aan 0.225. [8]

Wanneer men bij het kwantificeren van produktiviteit alleen het aantal soorten V telt, beperkt men zich tot een ééndimensionele analyse die in problemen komt bij een produktief suffix als *-sel* (0.080, 45), waarvan de typefrekwentie lager is dan dat van de improduktieve categorie van nomina actionis met

vocalische alternantie, in figuur 1 opgevoerd als VOCALT (0.004, 100). Betrekken we echter de kans op nieuwe soorten in de analyse, dan ontstaat een meer gedifferentieerd beeld, met het diminutiefsuffix en de N + N samenstellingen enerzijds en *-te* (0.013, 39) en *-nis* (0.013, 28) anderzijds als extrema. Ten aanzien van het suffix *-sel* kan men op de horizontale as aflezen dat het een veel grotere graad van produktiviteit bezit dan de categorie van nomina actionis met vocalische alternantie, ondanks het feit dat het aantal soorten beduidend lager ligt. Omgekeerd blijkt dat de graad van produktiviteit van *-te* beduidend lager is dan die van *-sel* ondanks het feit dat we voor beide categorieën nagenoeg hetzelfde aantal soorten tellen.

Interessant is ook de positie van het suffix *-ster* (0.231, 30), dat weliswaar een zeer lage pragmatische bruikbaarheid geniet, maar desondanks een hoge graad van produktiviteit: het is in het hedendaags Nederlands wel



Figuur 1. Pragmatische bruikbaarheid en graad van produktiviteit: nomina.

degelijk als levend affix beschikbaar. Als Brouwer (1984: 11) erop wijst dat ‘er over vrouwelijke vormingen wordt geredetwist waarbij bewuste keuzes plaatsvinden’, en dit aanvoert als een bewijs van een geringere produktiviteit van *-ster* dan in het algemeen wordt aangenomen, dan is het mijns inziens niet zozeer de graad van produktiviteit die ter discussie staat - op zich laten formaties in *-ster* zich met groot gemak vormen - maar de wenselijkheid van het gebruik van deze formaties, met andere woorden, hun pragmatische bruikbaarheid. Deze pragmatische bruikbaarheid is inderdaad zeer gering, maar lijkt de graad van produktiviteit, althans in het taalgebruik uit de zeventiger jaren, niet principieel aan te tasten. Zoals opgemerkt door Van Santen en de Vries (1981: 121), is er sprake van een ‘pragmatische factor van geringe bruikbaarheid in conceptueel opzicht’ in het spel die hier ‘remmend werkt’ op het aantal soorten V van het gemarkeerde lid van de oppositie *-er*: *-ster*. (Het totale aantal types van ‘personal agent nouns’ in *-er* of *-aar* bedraagt 312, voor *-ster* (het type *huurster*, 28 soorten, en de vormingen *leugenaarster* en *tuinierster*) tellen we slechts 30 soorten, hetgeen het grote verschil in pragmatische bruikbaarheid illustreert).

Op zichzelf genomen biedt het begrip ‘graad van produktiviteit’ geen zicht op het door Schultink (1962) verdedigde absolute verschil tussen produktiviteit en improduktiviteit, in de zin dat het begrip graad van produktiviteit alleen bij produktieve categorieën zinvol gehanteerd kan worden. Om de vraag te beantwoorden of een categorie produktief dan wel improduktief is op grond van onze kwantitatieve gegevens vergelijken we de categorie in kwestie met de verzameling ongelede woorden van dezelfde woordsoort. Immers, van de verzameling ongelede nomina, met N omschreven in figuur 1, weten we met zekerheid dat er van enige morfologische produktiviteit geen sprake kan zijn. Door te toetsen op de graad van produktiviteit [9] kan men nagaan of er van een significant verschil in de graad van produktiviteit sprake is. Bij uitvoering van de betreffende toetsen blijkt dat bijvoorbeeld de suffixen *-te* en *-nis* een graad van produktiviteit kennen die niet significant verschilt van die van de verzameling ongelede nomina. In het geval van nomina actionis met vocalische alternantie constateren we zelfs een significant lagere graad van produktiviteit. Met andere woorden, per morfologische categorie kan de vraag naar de absolute produktiviteit in principe [10] beantwoord worden.

Inmiddels kan de in de inleiding gestelde vraag naar de produktiviteit van *-te* worden beslecht in het voordeel van Schultink (1962). De graad van productiviteit van *-te* verschilt niet significant van die van de verzameling ongelede nomina. Bij *-heid* is daarentegen wel een significant verschil te vinden. Ook kan worden vastgesteld dat de graad van produktiviteit van *-heid* die van *-te* significant overtreft. [11]

Ten aanzien van het suffix *-atie* kunnen we opmerken dat voor beide suffixen *P* significant groter is dan de *P* waarde van de verzameling ongelede nomina. Voorts kan worden vastgesteld dat de graad van produktiviteit van *-ering* significant groter is dan die van *-atie* (0.10 versus 0.05), maar dat de pragmatische bruikbaarheid van *-atie* groter is dan die van *-ering* (184 versus 83). [12] Om het beeld van de produktiviteit van beide affixen te completeren is echter nadere informatie over de grootte van de onderliggende populaties

nodig. We komen derhalve in paragraaf 5 op deze kwestie terug.

4.2 Verba

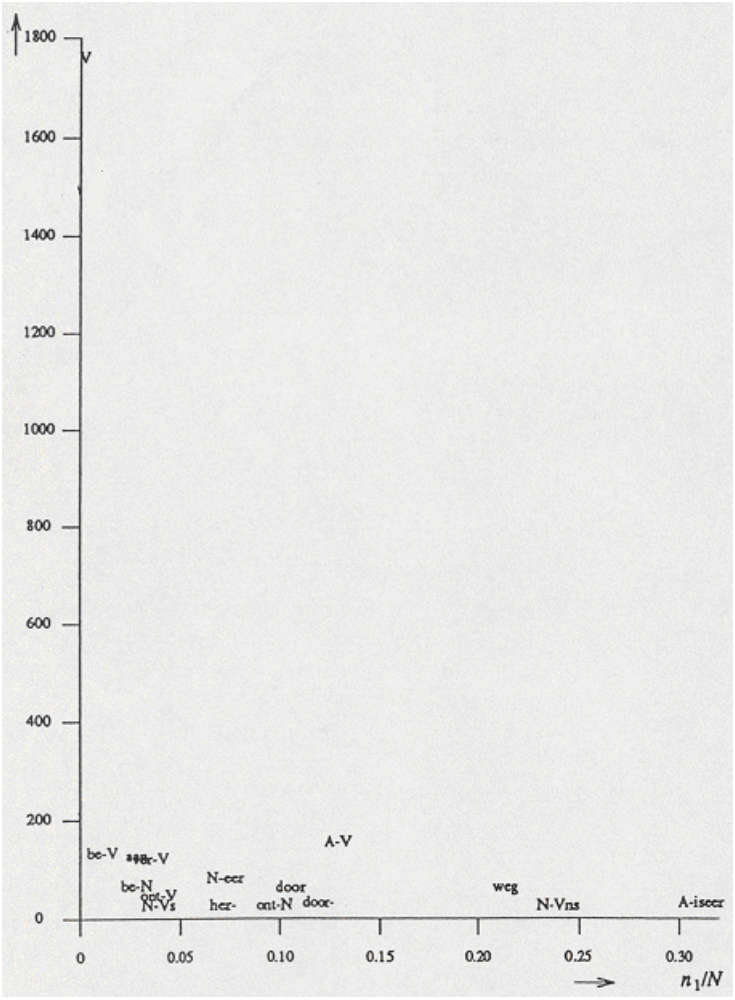
De verbale morfologische categorieën bieden een beeld dat nogal verschilt van dat van de nominale categorieën. Als zodanig bieden ze de mogelijkheid de in Baayen (1989) ontwikkelde theorie te toetsen. In figuur 2 vindt men in de linker bovenhoek de verzameling ongelede verba, aangeduid met V. Zij kenmerkt zich ten opzichte van de verzameling ongelede nomina door een nog geringere waarde van *P* en een groter aantal soorten. Voorts zijn opgenomen de prefixen *ver-*, *be-*, *her-* en *ont-*, de (scheidbare) partikels *aan*, *door* en *weg*, het suffix *-eer*, en verbale samenstellingen bestaande uit N + V (NV in figuur 2) of A + V (AV in figuur 2). De algemene indruk is dat zich bij de verbale categorieën vaker de situatie voordoet die we bij de nominale categorieën hebben aangetroffen voor *-ster* en *-sel*, namelijk een geringe pragmatische bruikbaarheid gecombineerd met een toch niet geheel te verwaarlozen graad van produktiviteit. De combinatie van een grote graad van produktiviteit en een grote soortenrijkdom, karakteristiek voor bijvoorbeeld *-heid*, *-tje* en de N + N samenstellingen, doet zich bij de verba niet voor. Anderzijds bevinden een aantal door de Vries (1975) als produktief gekarakteriseerde verbale categorieën zich in de linker benedenhoek van figuur 2, waar we juist de improduktieve categorieën verwachten. In vergelijking met de nominale categorieën in *-te* en *nis* valt echter op dat òf de graad van produktiviteit, òf het aantal soorten bij deze verbale categorieën groter is. Mede dankzij het feit dat de *P* waarde van de ongelede verba geringer is dan die van de ongelede nomina, blijkt bij toetsing de graad van produktiviteit van de produktieve gelede verbale categorieën toch nog (net) significant groter dan die van de ongelede verba. [13]

Na dit globale overzicht illustreren we de mogelijkheden van de kwantitatieve analyse in meer detail aan de hand van de verschillende in figuur 2 opgenomen categorieën. Wat betreft de partikels *door* en *weg* kunnen we kort zijn. De Vries (1985: 71-72, 83) acht samenkoppelingen met *door* (0.105, 60) en *weg* (0.214, 62) produktief, een oordeel dat in overeenstemming is met de gemeten waarden van *P* en *U*. Ook het prefix *-door* (0.119, 30) is onmiskenbaar produktief (vergelijk de Vries 1975: 137).

Het prefix *be-* hecht zich aan adjectiva (0.021, 21), verba (0.011, 130) en nomina (0.028, 64). Volgens de Vries (1975: 179) en Schultink (1962: 257) is de *be* + A categorie improduktief. Nu heeft het quotiënt $n1/N$ bij deze categorie een waarde die op zichzelf genomen niet onderdoet voor die van de *be* + N of *be* + V categorieën, hetgeen produktiviteit suggereert. Ondanks het feit dat de aangetroffen aantallen te gering zijn om een significantietoets uit te kunnen voeren blijkt toch uit de rest van de frekwentiegegevens dat het hier om een improduktieve categorie gaat: het aantal soorten met gebruiksfrekwentie 2 (n_2) is namelijk gelijk aan dat van het aantal met gebruiksfrekwentie 1 (n_1). Bij produktieve categorieën daarentegen is $n_1 >> n_2$. Als $n_2 \geq n_1$, dan duidt dit op een eindige populatie waarvan het aantal soorten door de steekproef voor een groot deel is uitgeput. Bij het vergroten van de steekproef neemt n_1 snel af, terwijl n_2 toeneemt. We

concluderen dat de *be* + A categorie kwantitatief gezien wel degelijk improduktief is.

De Vries (1975: 116, 171) beschouwt deverbaal en denominaal *be-* als produktief. Formaties die semantisch niet compositioneel zijn, bijvoorbeeld *bedrijven* of *beheren*, rekent hij terecht niet tot deze categorieën. Interessant is dat wanneer we de niet compositionele formaties wel in onze berekeningen zouden betrekken, we een reductie in de graad van produktiviteit zouden



Figuur 2. Pragmatische bruikbaarheid en graad van produktiviteit: verba.

bewerkstelligen. [14] Dit is een belangrijke eigenschap die elke zinvolle produktiviteitsmaat dient te bezitten. Immers, naarmate meer woorden semantisch niet meer compositioneel zijn, wordt de semantiek van de woordvormingsregel ondoorzichtiger en zal de regel aan produktiviteit inboeten (vergelijk Aronoff 1976, van Marle 1988). [15]

Ten aanzien van het prefix *her-* (0.072, 24) merkt de Vries op dat het produktief is, zij het dat deze produktiviteit 'in het gangbare Nederlands [is] beperkt tot het verzorgde geschreven Nederlands van de dag- en weekbladpers, en de wetenschappelijke verhandeling' (1975: 132). In het licht van deze observatie is de geringe pragmatische bruikbaarheid van *her-* - we tellen slechts 24 soorten - begrijpelijk. Anderzijds bezit het suffix een redelijke graad van produktiviteit, die de welgevormdheid van formaties als *herherverdelen* of *herheranalyseren* garandeert.

Combinaties van adjectief en verbum van het type *hardlopen, droogvallen, goedkeuren* (0.129, 152) scoren goed wat betreft pragmatische bruikbaarheid en graad van produktiviteit. Kwantitatief gezien is de karakterisering 'uitbreidbaar' (de Vries 1975: 94) dan ook conform de verwachting. Ten aanzien van combinaties van nomen en verbum beperken we ons tot die gevallen die niet door impliciete transpositie of backformation (het type *stofzuigen*) tot stand zijn gekomen (vergelijk de Vries 1975, hoofdstuk 4). Voor de categorie als geheel vinden we $P = 0.105$ en $U = 47$. Nadere uitsplitsing naar scheidbaarheid (*schoolzwemmen* (0.039, 24), NVns in figuur 2) versus *ademhalen* (0.240, 23), NVs in figuur 2) laat zien dat de verzameling niet scheidbare formaties nauwelijks uitbreidbaar is. Niet alleen is de graad van produktiviteit hier beduidend lager, maar ook vinden we dat het aantal dislegomena het aantal hapaxen overtreft (9 versus 4). De overstap van niet-scheidbaar naar scheidbaar - vergelijk de door de Vries (1975: 96-97) besproken vormen van het type *rijdt zwart* en *loopt hard-* wordt klaarblijkelijk makkelijker met het toenemen van de gebruiksfrekwentie van de formaties in kwestie. Dit ligt in de lijn der verwachting: naarmate activiteiten als *hardlopen* of *skilopen* vaker worden beoefend zal de wenselijkheid van de beschikbaarheid van finiete verbale vormen toenemen.

Het partikel *aan* leidt in de analyse van de Vries (1975: 75-76) tot drie onderscheiden typen samenkoppelingen, die elk op hun produktiviteit worden beoordeeld. Het type *aanraken* (0.021, 108), met een semantisch moment van nabijheid of oppervlakkig contact, wordt produktief geacht. Het type *aanbraden* (0.079, 12), met een ingressief semantisch moment, en het type *aanmodderen* (0.667, 5), met een duratief semantisch moment, zouden daarentegen improduktief zijn. Kwantitatief gezien doet zich echter het probleem voor dat met het afsplitsen van de laatste twee typen de graad van produktiviteit van het eerste, produktieve type lager wordt dan die van het totaal van alle formaties met het partikel *aan* (0.028, 125). Het verschil met de afsplitsing van een echt onregelmatige groep, boven geïllustreerd aan de hand van het deverbale *be-*, is opvallend. Dit suggereert dat de afsplitsing van de typen twee en drie, die trouwens per stuk een hogere *P* waarde te zien geven dan het eerste type, artificieel is. Mijns inziens is bij het partikel *aan* geen sprake van polyfunctionaliteit. De drie bovengenoemde semantische momenten liggen in eikaars verlengde en

kunnen als het resultaat van éénzelfde semantische functie op variërende semantische waarden van de grondwoorden worden begrepen. De grondbetekenis van *aan* laat zich mijns inziens omschrijven als een perfectief betekenismoment dat zich concretiseert als het bereiken van het met de door het grondwoord uitgedrukte handeling beoogde resultaat. Bij verba van beweging of verplaatsing (*aanlanden, aanraken*) duidt *aan* op het bereiken van het eindpunt van de 'path' (Talmy 1985), met als gevolg dat nabijheid of oppervlakkig contact wordt gerealiseerd. Het partikel heeft hier perfectieve waarde, geconcretiseerd als het bereiken van de met de beweging beoogde positie. Bij handelingsverba als *doen, zetten, drijven* vestigt *aan* de aandacht op het resultaat bereikt bij het einde van een handeling die een nieuwe toestand effectueert. Het ingressieve moment dat de Vries (1975) bij deze formaties signaleert is wederom een verbijzondering van de (perfectieve) grondbetekenis van *aan*. Bij het type *aanmodderen* is de door de Vries (1975: 76) gesignaleerde verplichte aanwezigheid van *maar* relevant. In een zin als

(3)
Hij moddert maar wat aan

relativeert *maar* het bereiken van het met de handeling beoogde resultaat, zodat een duratief betekenismoment wordt verkregen. In een zin als

(4)
Hij komt weer met zo'n verhaal aanzetten

heft *komt* het perfectieve aspect van *aan* op.

Deze heranalyse maakt tevens inzichtelijk waarom een werkwoord als *aanbraden*, dat volgens de Vries (1975: 75) een ingressief semantisch moment heeft, hiermee niet bevredigend is gekarakteriseerd. Bij *aanbraden* gaat het immers niet om een louter 'beginnen met braden', maar om een kortstondig proces van braden dat voorafgaat aan andere bereidingstechnieken. Met andere woorden, bij *aanbraden* is het perfectieve aspect, namelijk dat het beoogde resultaat snel bereikt is, minstens even relevant als het ingressieve moment.

Deze benadering is verwant met de door Botha (1988) voorgestelde analyse van reduplicatie in het Afrikaans, en sluit nauw aan bij de door Booij (1986) voorgestelde semantische analyse van het suffix *-er*. Booij (1986: 505-507) wijst op een drietal verklaringsmogelijkheden voor de polysemie van afgeleide woorden: een verklaring op grond van een algemene grondbetekenis geassocieerd met de woordvormingsregel, een verklaring op grond van een prototypische betekenis van de regel, en een verklaring waarbij gewezen wordt op de thematische rollen van de argumenten van het grondwoord. De eerste twee verklaringsmogelijkheden zijn in dit kader relevant. In het geval van het suffix *-er* stelt Booij (1986) dat er geen sprake is van polyfunctionaliteit maar van polysemie, in de zin dat de prototypische interpretatie 'personal agent' verbijzonderd kan worden tot 'impersonal agent' en zelfs tot 'instrument'. Deze analyse is ook vanuit kwantitatief oogpunt wenselijk. De volledige morfologische categorie van nomina in *-er* heeft als coordinaten (0.076, 299), bij uitsplitsing verschijnen voor de drie typen

'personal agent', 'impersonal agent' en 'instrument' de respectievelijke coördinaten (0.070, 267), (0.125, 18) en (0.778, 8). Wederom zien we dat een uitsplitsing in drie (of meer) semantisch disjuncte morfologische categorieën kwantitatief gezien contraproduktief is: de graad van produktiviteit van het hoofdtype neemt af, en is kleiner ten opzichte van de graad van produktiviteit van de neventypes. Een verschil tussen deze analyse van *-er* en die van *aan* is dat bij *-er* de prototypische interpretatie mogelijk blijft, ook wanneer de minder prototypische interpretatie meer gebruikelijk is (neem bijvoorbeeld het naamwoord *blaffer*. Bij de analyse van *aan* neem ik aan dat er sprake is van één grondbetekenis die nader geconcretiseerd wordt door kontekst, situatie en/of kennis van de wereld, in interactie met de betekenis van het grondwoord.

Het prefix *ver-* hecht zich aan adjectiva (0.026, 102), nomina (0.053, 63) en verba (0.035, 118). In elk van deze gevallen is het prefix produktief (de Vries 1975: 173, 178, 122). Ik beperk me hier tot het deverbale *ver-*, en breng een onderscheid aan tussen de formaties waar een negatief betekenismoment aanwezig is (0.035, 118), de formaties waar een dergelijk betekenismoment ontbreekt (0.009, 29), en de formaties die semantisch niet als compositioneel zijn te beschouwen (0.005, 43). De eerste, produktieve categorie, die een aantal categorieën van de Vries (1975) samenneemt, omvat formaties als *verbranden*, *vermoorden*, *verkoken* en *verhongeren*, verba die handelingen tot uitdrukking brengen die een negatief waardeerbare verandering uitwerken op het direct object. Formaties waarbij het veroorzaken van een dergelijke negatieve uitwerking moeilijker te denken valt, zoals bijvoorbeeld *slikken* of *rekenen*, prefigeren *ver-* met name in combinatie met het reflexieve *zich* (de Vries 1975: 122). Het gebruik van *zich* wordt hier door de semantiek van het grondwoord afgedwongen. Gegeven de grondbetekenis van *ver-*, 'het door de met het grondwoord gegeven handeling teweegbrengen van een negatief effect op het object van dit grondwoord' [16], ontstaat bij verba als *verspellen* en *verslikken* het probleem dat men door te spellen of te slikken moeilijk een negatief effect op een object anders dan het subject kan bewerkstelligen. Door het gebruik van *zich* wordt de vereiste reflexieve interpretatie verkregen, en is een zinvol gebruik van *ver-* gegarandeerd. [17] Bij constructieve werkwoorden als *bouwen*, *naaien* en *smeden* drukt *ver-* een positief te waarderen verandering uit die echter via een aanvankelijk negatief effect op het object, namelijk een gedeeltelijke vernietiging, tot stand komt. Tenslotte is ook bij verba als *verjagen* en *vervloeken* een (bedoeld) negatief betekenismoment aanwijsbaar, vergelijk *vloeken* met *vervloeken*.

Terwijl de Vries (1975: 119-126) een aantal verschillende categorieën onderscheidt en daarvan slechts de typen *vermoorden* (0.047, 43) en *zich verslikken* (0.013, 33) produktief acht, is in de hier voorgestelde heranalyse sprake van één enkele produktieve categorie. Kwantitatief gezien heeft deze analyse als voordeel dat de marginale produktiviteit van het type *zich verslikken* niet langer problematisch is. Het produktiviteitsoordeel van de Vries (1975) ten aanzien van dit type kan nu begrepen worden tegen de achtergrond van het feit dat het deel uitmaakt van een grotere, overkoepelende categorie die kwantitatief gezien veel sterker staat.

Als laatste prefix analyseren we *ont-*. Volgens de Vries (1975: 172) mag men bij het denominale *ont-* 'tot een zekere produktiviteit besluiten'. Dit voorzichtige oordeel wordt bevestigd door het zeer geringe aantal geconstateerde soorten (24), dat in eerste instantie improduktiviteit doet vermoeden, en de niet onredelijke graad van produktiviteit (0.097), die een bijstelling van deze aanvankelijke inschatting afdwingt.

Het deverbale *ont-* zou volgens de Vries (1975: 127-129) improduktief zijn. Hij komt tot deze conclusie tegen de achtergrond van de zijns inziens zeer verschillende wijzen waarop dit prefix de betekenis van het grondwoord modificeert. Hij onderscheidt een drietal categorieën, te weten een inchoatieve categorie (type *ontbranden*, (0.020, 9)), een categorie van het type *ontbinden* (0.029, 15), en een separatieve categorie (het type *ontsnappen*, (0.073, 20)). Geen van deze categorieën is zijns inziens produktief.

Vanuit een iets gewijzigde optiek is de semantiek van het deverbale *ont-* homogener dan deze analyse doet vermoeden. Als we aannemen dat *ont-* als grondbetekenis heeft 'een verandering in tegengestelde richting teweeg brengen', blijken de door de Vries onderscheiden categorieën onder één noemer gebracht te kunnen worden. Bij werkwoorden als *ontsluiten*, *onttoveren*, *ontladen* geeft *ont-* aan dat de door het grondwoord uitgedrukte handeling met tegengesteld doel of effect wordt verricht. Bij verba als *ontstelen*, *ontlokken*, *ontvoeren* wordt de met het grondwoord gegeven handeling voltrokken met een intentie tegengesteld aan die van het object (*ontvoeren*) of indirect object (*ontlokken*). Bij *ontvluchten, ontlopen* en *ontduiken* is het betekenismoment van de tegengestelde intentie eveneens essentieel. Kenmerkend voor deze verba is niet zozeer een separatief betekenismoment - vergelijk *wegvluchten* en *weglopen* - als wel het feit dat de handeling uitgedrukt door het grondwoord plaatsvindt met een intentie die tegengesteld is aan die van anderen.

Bij irreversibele processen als (*branden*, *ploffen*, *kiemen*, *groeien*) is 'het teweegbrengen van een verandering in tegengestelde richting' in letterlijke zin conceptueel problematisch. Met een inchoatieve interpretatie wordt echter toch aan de grondbetekenis van *ont-* tegemoetgekomen, in de zin dat het contrast tussen de toestand voor en na het intreden van het proces wordt belicht. Dat dit contrast door *ont* als relevant wordt opgepakt blijkt bij vergelijking met de perifrastische inchoatieve constructies *gaan groeien*, *gaan kiemen*, *gaan branden* en *gaan ploffen*.

Is deze heranalyse juist, dan hebben we te maken met één morfologische categorie met 44 soorten en een produktiviteitsgraad van 0.039. Deze waarden maken het onwaarschijnlijk dat we met een volledig improduktieve categorie van doen hebben. We zien ons dus geplaatst voor een discrepantie tussen de kwantitatieve analyse en het produktiviteitsoordeel van de Vries (1975). Nu is het echter de vraag of dit produktiviteitsoordeel volledig juist is. Mijn twijfel is op de volgende drie overwegingen gebaseerd. In de eerste plaats werkt een fragmenterende semantische analyse een negatief produktiviteitsoordeel in de hand: men ziet zich hierdoor geplaatst voor een aantal kleine categorieën die stuk voor stuk, dankzij hun specifieke semantische definitie, minder goed uitbreidbaar zijn. In de tweede plaats zijn nieuw-

vormingen hier niet principieel onmogelijk. Een neologisme als *ontfriemelen* is mijns inziens wel degelijk grammatikaal. In de derde plaats beschouwt de Vries (1975: 172) formaties als *onterven*, *ontgrendelen*, *ontkoppelen*, *ontstemmen* en *ontwortelen* als afgeleid van *erf*, *grendel*, *koppel*, *stem* en *wortel* in plaats van van de verba *erven*, *grendelen*, *koppelen*, *stemmen* en *wortelen*. Op grond van de veronderstelde improduktiviteit van het deverbale *ont-* neemt hij tevens aan dat *ontkleden* direct van *kleed* is afgeleid en niet van *kleden*. Deze analyse is onjuist. De echt denominale formaties, bijvoorbeeld *ontbossen*, *ontkrachten*, *ontvolken* en *ontkurken* laten zich verstaan als een vrijwel letterlijk 'van bos, kracht, volk en kurk ontdoen'. Deze interpretatie is echter niet van toepassing op *onterven*, *ontgrendelen* of *ontkleden*. Wie zich ontkleedt ontdoet zich niet van een kleed maar van zijn kleren, hij voert het proces van zich kleden in omgekeerde richting uit. Als een clavecimbel ontstemt wordt deze niet van zijn stem ontdaan, maar verslechtert de aangebrachte stemming. Wie wordt onterfd ontloopt zijn erfenis maar behoudt zijn erf. In al deze gevallen heeft het (ongelede) werkwoord een meerwaarde ten opzichte van het gelijkvormige naamwoord, en deze meerwaarde vinden we terug bij de formaties met *ont-*, hetgeen wijst op een deverbale, niet een denominale oorsprong. Nu beschouwt de Vries (1975) formaties van het type *onterven* als produktief. Aannemende dat dit produktiviteitsoordeel onafhankelijk is van de door hem veronderstelde derivationele herkomst, kunnen we concluderen dat in het minimale geval de produktiviteit van het deverbale *ont-* een open vraag is. In het licht van de vermelde kwantitatieve bevindingen waag ik me echter aan de hypothese van een wellicht geringe, doch niet verwaarloosbare graad van produktiviteit.

Tenslotte rest ons de analyse van het suffix *-eer*. Volgens de Vries (1975: 180-181) gaat het hier om een perifeer, improduktief proces dat zijn oorsprong vindt in taalcontact en van daaruit mogelijkerwijs ondersteund wordt. Zonder een ondersteunende rol van Engelse en Franse morfologische procédés te willen ontkennen, moet toch opgemerkt worden dat als resultaat van dit taalcontact het Nederlands, althans in geschreven taal, beschikt over een produktief denominaal *-eer* (0.078, 70) en de-adjectivaal *-iseer* (0.283, 29) [18], categorieën die gunstig afsteken bij veel van de in figuur 2 opgenomen zuiver inheemse categorieën.

5. Pragmatische potentialiteit

Zoals boven uiteengezet gaat het bij pragmatische potentialiteit om de vraag in hoeverre het aantal soorten in de steekproef het aantal soorten van de populatie uitput. Bij improduktieve categorieën verwachten we dat S/V klein is en in de limiet voor $N \rightarrow \infty$ naar 1 gaat, bij produktieve categorieën is dit quotiënt of oneindig, of groot. Tabel 3 geeft voor een aantal categorieën het aantal geconstateerde soorten V, de geschatte grootte van de populatie S en het quotiënt $I = S/V$. Uit deze tabel blijkt dat ook de produktiviteitsmaat I een plausibele ordening naar groeiende produktiviteit aanbrengt. Voor het zeer produktieve diminutiefsuffix is I inderdaad oneindig, voor de improduktieve suffixen *-nis* en *-te* vinden we de laagste waarden van I.

Ondanks het feit dat het aantal mogelijke woorden in *-heid* niet oneindig is, [19] valt het verschil met de lage waarde van het rivaliserende suffix *-te* op. De stelling dat *-te* improduktief is wordt hiermee eens te meer bevestigd. We zijn nu ook in staat de produktiviteit van *-atie* en *-ering* volledig in te schatten. Intuïties die terug te voeren zijn op de graad van produktiviteit laten een duidelijk verschil zien tussen *-atie* en *-ering*, in het voordeel van het laatste suffix. Met andere woorden, *-atie* is minder produktief dan *-ering*, zonder dat het als improduktief kan worden beschouwd. Onderzoeken we de potentialiteit van de beide categorieën, dan valt op dat het geschatte aantal mogelijke soorten in beide gevallen laag uitvalt in vergelijking met bijvoorbeeld de aantallen formaties eindigend in de suffixen *-heid, -er* of *-je*. Twijfels aangaande de algehele produktiviteit van de niet inheemse suffixen *-atie* en *-ering* zijn vermoedelijk terug te voeren op dit geringe aantal mogelijke soorten. Improduktief zijn ze echter niet: noch *P*, noch *I* zijn uitzonderlijk laag in vergelijking met bijvoorbeeld een produktieve inheemse categorie als *-sel*.

Tabel 3. Pragmatische potentialiteit

affix	***V***	***S***	***I***
-nis	23	32	1.4
-te	39	59	1.5
-atie	186	297	1.6
-ing	943	1612	1.7
be + N	64	121	1.9
door	60	161	2.7
-sel	45	126	2.8
weg	62	194	3.1
-ering	83	331	4.0
-heid	466	2063	4.4
-tje	1031	∞	∞
-er	299	∞	∞

In het licht van deze analyse wordt het duidelijk waarom de door Al & Booij (1981) voorgestelde analyse niet tot het gewenste resultaat kan leiden. Door alleen het aantal soorten *V* in de berekening te betrekken wordt voorbijgegaan aan de potentialiteit van woordvormingsregels. Door gebruik te maken van het volledige frekwentiespectrum kunnen we met behulp van *P* en *I* deze potentialiteit in kaart brengen en een genuanceerder analyse ontwikkelen voor de productiviteit van rivaliserende affixen als *-atie*, *-ering* en *-heid*, *-te*.

Tenslotte moet nog gewezen worden op de relatief geringe waarden van *I* voor de verbale categorieën. Dit feit staat in verband met het bijna volledig ontbreken van affixstapeling bij verba. Uit de in de appendix opgenomen waarden van het gemiddelde aantal morfemen in het gelede woord blijkt dat, met uitzondering van *be* + A, de verbaliserende prefixen een zeer duidelijke voorkeur tonen voor een monomorfematisch grondwoord. Formaties van het type *herbebossen* en *veraanschouwelijken* zijn eerder uitzondering dan regel. Echter, produktieve processen manifesteren hun

produktiviteit en potentialiteit met name in hun meervoudig gelede formaties. Zo heeft het suffix *-heid* een gemiddelde van 3.31 morfemen per formatie. Het feit dat gelede verba juist hier ten zeerste worden beperkt heeft tot gevolg dat de aanname van een aftelbaar oneindige onderliggende populatie irrealistisch wordt, en door de statistische analyse wordt geloochenstraft. Immers, om 'de aanvoer van nieuwe basiswoorden' (Uhlenbeck 1953: 336) te garanderen is het noodzakelijk dat er, naast het (eindige) bestand aan ongelede woorden, een in principe aftelbaar oneindig bestand aan gelede woorden beschikbaar is. Juist bij verba, waar van recursie nauwelijks sprake is, wordt dit problematisch. De reden van de zeer beperkte mogelijkheden van affixstapeling moeten wellicht gezocht worden in het optreden van conceptualiseringsproblemen bij voortschrijdende deverbale derivatie. Wat de uiteindelijke reden van dit verschijnsel ook mag blijken te zijn, als zodanig vormt het een factor die op zijn beurt de produktiviteit beperkt van het suffix *-ing*, dat verhoudingsgewijs zelden aan grondwoorden met meer dan twee morfemen wordt aangetroffen. [20]

6. Slotopmerkingen

De hier gegeven schets van de mogelijkheden van een kwantitatieve benadering van het verschijnsel produktiviteit is verre van volledig. Zo heb ik de vraag naar de representatie van voor token-frekwentie gevoelige woordvormingsregels niet ter sprake gebracht. Ook op de psycholinguïstische interpretatie van onze bevindingen kan in dit kader niet nader worden ingegaan. Voor nadere beschouwingen over deze kwesties raadplege men Baayen (1989).

Dat een statistische analyse nieuwe mogelijkheden biedt bij het onderzoek van het fenomeen van morfologische produktiviteit meen ik voldoende duidelijk te hebben gemaakt. De kwantitatieve methode is vooral van waarde voor de minder duidelijke gevallen, waar door een complexe interactie van heterogene factoren de herkomst van de intuïties slecht traceerbaar is. Zij kan tevens als een heuristische techniek die complementair is ten opzichte van een op intuïties gebaseerde analyse in het morfologisch onderzoek worden gehanteerd.

Tenslotte biedt deze methode een goed uitgangspunt voor het onderzoek van morfologische produktiviteit in dode talen als Latijn, Klassiek Grieks of Klassiek Hebreeuws, talen waarvoor de intuïties van 'native speakers' niet kunnen worden geraadpleegd.

Appendix

categorie	*N*	*V*	n_1	n_2	*P*	$\hat{\sigma}_{n1}$	*Z*	*m*	σ_m
door	257	60	27	9	0.105	3.56	7.46	2.05	0.22
door +	109	30	13	9	0.119	2.88	4.45	2.00	0.00
weg	154	62	33	11	0.214	5.17	6.33	2.00	0.00
be + A	193	21	4	4!	0.021	-	-	2.52	0.60
be + V	2265	130	25	14	0.011	1.87	11.28	2.02	0.15
be + N	725	64	20	9	0.028	2.40	7.82	2.08	0.27
A + V	520	152	67	30	0.129	5.99	11.04	2.01	0.08
N + V	152	47	16	13	0.105	-	-	2.00	0.00
N + V, s	102	24	4	9!	0.039	-	-	2.00	0.00
N + V, ns	50	23	12	4	0.240	-	-	2.00	0.00
aan	1192	125	33	16	0.028	2.60	11.91	2.00	0.00
ver + A	784	102	20	19	0.026	1.94	9.62	2.16	0.39
ver + N	416	63	22	19	0.053	1.85	11.51	2.03	0.25
ver + V	953	118	33	17	0.035	4.58	6.85	2.00	0.00
ont + N	124	24	12	8	0.097	2.81	4.20	2.04	0.20
ont + V	331	44	13	7	0.039	3.08	4.04	2.00	0.00
N + eer	358	79	26	17	0.073	2.61	9.73	2.01	0.11
A + iseer	61	30	19	7	0.311	-	-	2.33	0.55
VO	169602	1757	290	180	0.002	8.08	-	-	-

m: gemiddeld aantal morfemen; VO: ongelede verba.

Bibliografie

Al, B.P.F. & Booij, G.E., 1981. 'De produktiviteit van woordvormingsregels. Enige kwantitatieve verkenningen op het gebied van nomina actionis'. *Forum der Letteren 22*, 26-38.
Anshen, F. & Aronoff, M., 1988. 'Producing Morphologically Complex Words'. *Linguistics 26*, 641-655.
Aronoff, M., 1976. *Word Formation in Generative Grammar*. Cambridge, Mass., MIT press.
Baayen, R.H., 1989. *A Corpus-Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation*. Diss. Vrije Universiteit, Amsterdam.
Bolinger, D.L., 1948. 'On defining the morpheme', in: Bolinger, D.L (ed.), *Forms of English, Accent, Morpheme, Order*, Cambridge, Mass., Harvard University Press, 183-189.
Booij, G.E., 1977. *Dutch Morphology. A Study of Word Formation in Generative Grammar*, Dordrecht, Foris.
Booij, G.E., 1986. 'Form and Meaning in Morphology: the Case of Dutch 'Agent Nouns'. *Linguistics* 24, 503-517.
Booij, G.E. & Haaften, T. van, 1988. 'The external syntax of derived words: evidence from Dutch. *Yearbook of Morphology 1*, 29-44.
Botha, R.P., 1988. *Form and meaning in word formation. A study of Afrikaans reduplication*. Cambridge, Cambridge University Press.
Brouwer, D., 1984. 'Hoe vrouwelijk is een hoofd'? Over de vorming van vrouwelijke beroepsnamen in het Nederlands, in: G.J. de Haan, M. Trommelen & W. Zonneveld (eds.). *Van Periferie naar Kern*. Foris, Dordrecht 11-27.
Burgschmidt, E., 1977. 'Strukturierung, Norm and Produktivität in der Wortbildung, in: Brekle, H.E. & D. Kastovsky (eds.), *Perspektiven der Wortbildungsforschung*, Bonn, Bonvier Verlag, 1977, 39-47.

Cedergren, H.J. & Sankoff D., 1977. 'Variable Rules: Performance as a Statistical Reflexion of Competence. *Language 50*, 333-355.
Dale, van, 1961[8]. *Groot Woordenboek der Nederlandse Taal (C. Kruyskamp)*. Den Haag, Nijhof.
Downing, P., 1977. 'On the Creation and Use of English Compound Nouns'. *Language 53*, 810-842.
Harwood, F.W. & Wright, A.M., 1956. 'Statistical Study of English Word Formation'. *Language 32*, 260-273.
Hoekstra, T., 1984. *Transitivity*. Dordrecht, Foris.
Marle, J. van, 1985. *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht, Foris.
Marle, J. van, 1988. 'Betekenis als Factor bij Productiviteitsverandering'. *Spektator 17-4*, 341-359.
Rainer, F., 1988. 'Towards a Theory of Blocking: the Case of Italian and German Quality Nouns'. *Yearbook of Morphology 1*, 155-185.
Santen, A. van & Vries, J.W. de, 1981. 'Vrouwelijke persoonsnamen op -*ster*. *Forum der Letteren 22*, 115-125.
Schultink, H., 1962. *De Morfologische Valentie van het Ongelede Adjectief in Modern Nederlands*. Den Haag, van Goor Zonen.
Talmy, L., 1985. 'Lexicalisation Patterns: Semantic Structure in Lexical Forms, in: T. Shopen (ed.), *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*. Cambridge, Cambridge University Press.
Thorndike, E.L., 1943. 'Derivation Ratios'. *Language 19*, 27-37.
Uhlenbeck, E.M., 1953. 'The Study of Word-classes in Javanese'. *Lingua 3*, 322-354.
Uit den Boogaart, P.C. (ed.), 1975. *Woordfrequenties in Gesproken en Geschreven Nederlands*. Utrecht, Oosthoek, Scheltema & Holkema.
Vries, J.W. de, 1975. *Lexicale Morfologie van het Werkwoord in Modern Nederlands*. Leiden, University Pers.

Eindnoten:

* Ik ben Geert Booij, Piet van Reenen en Ariane van Santen zeer erkentelijk voor hun commentaar op een eerdere versie van dit artikel.

1 Deze neologismen op -*erd* en -*heid* komen niet voor in de CELEX bestanden of in van Dale (1961).
2 Tellingen van aantallen soorten dienen op corpora en niet op woordenboeken gebaseerd te zijn, vergelijk Al & Booij (1981) en recentelijk Anshen & Aronoff (1988).
3 In tabel 1 duidt 0 op de afwezigheid van beperkende factoren, 1 op de voorwaarde dat in het grondwoord -*eer* direct op een lexicaal morfeem volgt, en 2 op de eis dat het grondwoord bisyllabisch is.
4 Het gaat hier om die adjectiva die ongeleed zijn en niet eindigen in een vokaal of in een konsonantkluster waarvan de eerste konsonant geen sonorant is.
5 Vergelijk Downing (1977)'s voorbeeld van de deictische samenstelling *apple juice chair* als aanduiding voor een stoel waar een glas appelsap bij staat.
6 Vergelijk Burgschmidt (1977) voor het onderscheid tussen grammatikaliteit en 'geschiktheid'.
7 De index *I* is een inverse variant van Aronoff (1976)'s index van produktiviteit.

8 In het vervolg vatten we de produktiviteit van een categorie samen met behulp van de coordinaten van de categorie in het (*P, U*) vlak. De frekwentiegegevens van de nominale categorieën, met uitzondering van de N + N samenstellingen, zijn ontleend aan Baayen (1989). Voor de N + N samenstellingen en de verbale categorieën is gebruik gemaakt van de CELEX bestanden.
9 Voor de te hanteren toets zie Baayen (1989), hoofdstuk 5.
10 Voorwaarde is wel dat met name de aantallen soorten met de laagste gebruiksfrekwenties voldoende groot zijn om de benodigde berekeningen uit te voeren.
11 Voor aanvullende kwantitatieve argumenten voor de improduktiviteit van -*te* zie Baayen (1989), hoofdstuk 9.
12 Voor statistische details zie Baayen (1989), hoofdstuk 10.
13 Zie de appendix voor de statistische details.
14 Voor *be* + *V* neemt de graad van produktiviteit af van 0.011 tot 0.008, voor *be* + *N* vinden we een reductie van 0.028 naar 0.017.
15 Merk op dat het mee in de berekening betrekken van semantisch en ook formeel idiosyncratische gelede woorden de relatieve toepassingswaarschijnlijkheid (Al & Booij 1981) van het improduktieve affix verhoogt in plaats van verlaagt. Aangezien alleen het aantal soorten in de analyse wordt betrokken, en gegeven dat we bij improduktieve categorieën meer onregelmatige vormen verwachten dan bij produktieve categorieën, zal een verschuiving ten gunste van het improduktieve affix optreden. Hiermee is de onbruikbaarheid van de methode van Al & Booij (1981) nog eens vanuit een andere invalshoek belicht.
16 Deze omschrijving is ook van toepassing op de intransitieve, niet-accusatieve lezingen van de verba *verkoken* en *verhongeren* onder de aanname dat deze verba geen extern argument hebben, maar alleen een direct intern argument (vergelijk Hoekstra 1984).
17 Zie Booij & van Haaften (1988) voor het effect van de semantiek van woordvormingsregels op de argumentstructuur van gelede woorden.
18 Formaties met stamallomorfie, bijvoorbeeld *hypnose* - *hypnotiseer*, zijn in deze tellingen niet opgenomen.
19 Het voor -*heid* voorspelde aantal soorten, 2063, is van de orde van grootte, 2399, van alle geregistreerde formaties in -*heid* in het Nederlands. Slechts 1792 van deze formaties komen tenminste tweemaal voor in het INL corpus (CELEX).
20 Een verfijnder analyse van de verhouding van het aantal morfemen en woordfrekwenties kan men vinden in Baayen 1989, hoofdstuk 6.